ITR08 Lecture 6

Valid XML

Thomas Krichel

2000–02–22

Reading

Young, Michael J, (2000) "Step-by-step XML", Microsoft Press, 2000

Bray, Tim, (1998) "The Annotated XML Specification", http://www.xml.com/axml/testaxml.htm (advanced)

Well-formed and valid documents

All XML documents must obey to certain constraint. An XML that does obey the constraints is called well-formed.

An XML document is called valid, if, in addition to being well-formed, it obeys some further constraints. These constraints are being called validity constraint.

Validity constraints are different from well-formedness constraints because they do not not apply to all XML documents but only to some.

DTD and Schema

A DTD is the classic way in which constraints on SGML documents can be defined. XML can be derived out of full SGML with the use of a DTD. XML can be further constraint through the use of DTDs.

XML Schemas are new ways in which constraints can be applied to XML documents. These constraints are much more refined. For example, they allow to constrain element contents, for example to say that the time of a track is of the form *1_or_2_digits*:*2_digits*.

There are different schema languages that are proposed on the web. These are experimental.

The sofix DTD part 1

A DTD contains element and attribute declarations. It is best leaned through the use of an example.

Look at the sofix DTD at

http://wotan.liu.edu/home/krichel/itr8/sofix.dtd

```
<!ELEMENT sofix (item+)>
```

means that there is an element `sofix` that has one or more children called `item`. One ore more is expressed by the `+`. `?` would mean 0 or 1 children, and `*` would mean zero or more children.

The sofix DTD part 2

The first line

```
<!ELEMENT item (labelname,number,work+)>
```

says that the element `item` accepts one element `labelname`, one element `number`, and several element `works` as children, in that order. If the comma (,) would be a vertical bar (|) it would mean that the item would contain either of these elements.

```
<!ELEMENT labelname (#PCDATA)>
```

says that the element labelname contains parsed character data.

The sofix DTD part 3

The next line declares that `%roles` is an entity. Whenever the `%roles;` (the semicolon says that the entity is finished there) then it is replaced with the contents of the file whose URL is given after the keyword system

```
<!ENTITY % roles SYSTEM
"http://wotan.liu.edu/home/krichel/itr8/sofix_roles">
```

This declaration is then used to say that the item `contributor` has required element `role` that takes its value of the list `%roles`.

```
<!ATTLIST contributor role %roles; #REQUIRED>
```

Making the document valid

1. Add the line

```
<!DOCTYPE sofix SYSTEM
"http://wotan.liu.edu/home/krichel/itr8/sofix.dtd">
```

before the root element, after the xml processing instruction.

2. Download

```
http://wotan.liu.edu/home/krichel/itr8/validitycheck.html
```

and edit to replace `file_to_check` with the name of your file.

3. Open `validitycheck.html` using the Internet explorer. Don't get angry.

to be done for next Tuesday

Email Thomas the checked valid sofix XML file containing at least three items.